K-사이버 시큐리티 챌린지 2019

AI 기반의 네트워크 위협탐지

와이앤슈

권영길, 이주현, 김지혜, 나예원, 이세은

0

목차

01

문제 파악, 데이터 분석

- 대회 문제
- 데이터 분석
- 데이터 전처리

02

피쳐 추출

- 피쳐 선정
- 피쳐 추출

03

비지도 학습 모델

- 이상 행위 탐지
- 공격 유형 분류

04

결론

- 본선 결과
- 활용 방안

1

문제 파악 및 데이터 분석

대회 문제

“ 대용량 악성/정상 트래픽을 분석하는

AI 기반의 네트워크 위협 탐지 알고리즘 개발 ”

가상의 네트워크인 A network와 실제 네트워크인 B network에서

정교화되고 증가하고 있는 네트워크 위협 행위를 자동으로 탐지하는

AI 알고리즘을 개발하고, **악성/정상 및 공격 유형을 탐지**하십시오

1

문제 파악 및 데이터 분석

데이터 분석

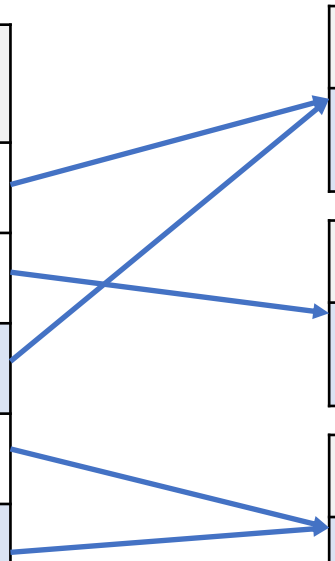
- 데이터 확인 결과 지도 학습으로는 학습이 불가능 하다고 판단
  - 정상과 악성 비율이 Unbalanced함
  - 모든 공격 유형에 대한 데이터가 없음

	학습 데이터셋(정상)			예선 데이터셋(정상+악성)		본선 데이터셋(정상+악성)	
	Train_set1 (Network A)	Train_set2 (Network B)	Brute force	Test1_1st (Network A)	Test1_2nd (Network B)	Test2_1st (Network A)	Test2_2nd (Network B)
사용한 Protocol 개수	54개	32개	2개	56개	58개	48개	77개
사용한 IP 개수 (src, dst)	약 9600 개	약 4000 개	2개	약 8400 개	약 1 만개	약 5500개	약 15000개
전체 통신의 개수	약 1170만 개	약 45 만 개	약 200개	약 930만 개	약 970 만개	약 500만개	약 2000만개

데이터 전처리

- 데이터 개수가 많은 관계로 통신 별(외부→내부 통신)로 데이터를 나눔
  - src.ip(외부) 와 dst.ip(내부) 통신을 기준으로 데이터를 분리
  - 정상/악성을 분류한 후 공격 유형을 파악하기 위해 Port 정보는 분리

src.ip	dst.ip
211.34.4.21	192.168.0.1
211.34.4.41	192.168.0.1
192.168.0.1	211.34.4.21
211.34.4.21	192.168.0.2
192.168.0.2	211.34.4.21



Src.ip	Dst.ip
211.34.4.21	192.168.0.1

Src.ip	Dst.ip
211.34.4.41	192.168.0.1

Src.ip	Dst.ip
211.34.4.21	192.168.0.2

2

피쳐 선정 및 추출

피쳐 선정

- 총 22개의 Feature 가 주어졌으나 아래 Feature를 이용해서는 학습이 불가
  - 프로토콜 마다 가지고 있는 값이 다름(TCP, UDP는 둘 중에 1개만 사용)
  - 데이터에 결측치(NULL) 값이 너무 많음(ARP, CDP의 경우 IP 값이 없음)
- 주어진 22개의 Feature 를 이용하여 새로운 Feature 생성 및 추출

Time	Protocol	IP	TCP	UDP	HTTP	
_ws.col.UTCtime	_ws.col.Protocol	ip.src	tcp.srcport	udp.srcport	http.request.method	http.response.code
		ip.dst	tcp.dstport	udp.dstport	http.request.uri	http.server
			tcp.len	udp.length	http.user_agent	http.content_type
			tcp.seq		http.connection	http.content_length
			tcp.ack		http.host	http.cache_control

2

피쳐 선정 및 추출

피쳐 추출

- CIC 에서 제시한 Feature 및 EDA를 통해 Feature 생성

No	Feature	Description
1	tot_fw_pk	통신한 총 패킷 개수
2	fw_iat_tot	전체 통신 시간
3	fw_iat_avg	이전 패킷과 다음 패킷 사이에 걸린 시간의 평균
4	fw_iat_std	이전 패킷과 다음 패킷 사이에 걸린 시간의 표준편차
5	fw_iat_max	이전 패킷과 다음 패킷 사이에 걸린 시간의 최대값
6	fw_iat_min	이전 패킷과 다음 패킷 사이에 걸린 시간의 최소값
7	fw_pkt_s	1초당 통신한 패킷의 개수
8	ratio	tcp.srcport와 tcp.dstport의 비율

2

피쳐 선정 및 추출

피쳐 추출

- fw_iat_avg : 이전 패킷과 다음 패킷 사이에 걸린 시간의 평균

	A	B	C	D	E	F
1	_ws.col.UTCtime	_ws.col.Protocol	ip.src	ip.dst	tcp.srcport	tcp.dstport
2	12:17:53	TCP	172.16.0.1	192.168.10.50	52108	21
3	12:17:56	FTP	172.16.0.1	192.168.10.50	52108	21
4	12:17:56	FTP	192.168.10.50	172.16.0.1	21	52108
5	12:17:57	FTP	172.16.0.1	192.168.10.50	52108	21
6	12:17:57	TCP	192.168.10.50	172.16.0.1	21	52108
7	12:17:57	FTP	192.168.10.50	172.16.0.1	21	52108
8	12:19:20	TCP	172.16.0.1	192.168.10.50	52112	21

  - ➢ 각각의 패킷을 시간으로 변환한 후, 각 패킷 사이의 시간의 평균을 계산

2

피쳐 선정 및 추출

피쳐 추출

- ratio : tcp.srcport와 tcp.dstport의 비율

tcp.srcport의 개수	tcp.dstport의 개수	ratio
1	100	0.01
10	10	1
100	1	100

➢ src/dst 의 포트를 중복제거를 하여 사용된 포트의 개수의 비율을 계산

2

피쳐 선정 및 추출

피쳐 추출

- 공격유형 탐지를 위한 Feature 추출

No	Feature	Description
9	is_in_bot	http.user_agent에 bot에 관련된 문자열이 포함되어 있는지 여부
10	Is_in_ares	http.server에 bot과 관련된 문자열이 포함되는지 여부
11	is_in_xss	http.request.uri에 XSS 공격과 관련된 문자열이 포함되어 있는지 여부
12	is_in_sql	httprequest.uri에 SQL injection 공격과 관련된 문자열이 포함되어 있는지 여부
13	check_rdp	tcp.dstport로 3389 포트를 사용하는지 여부

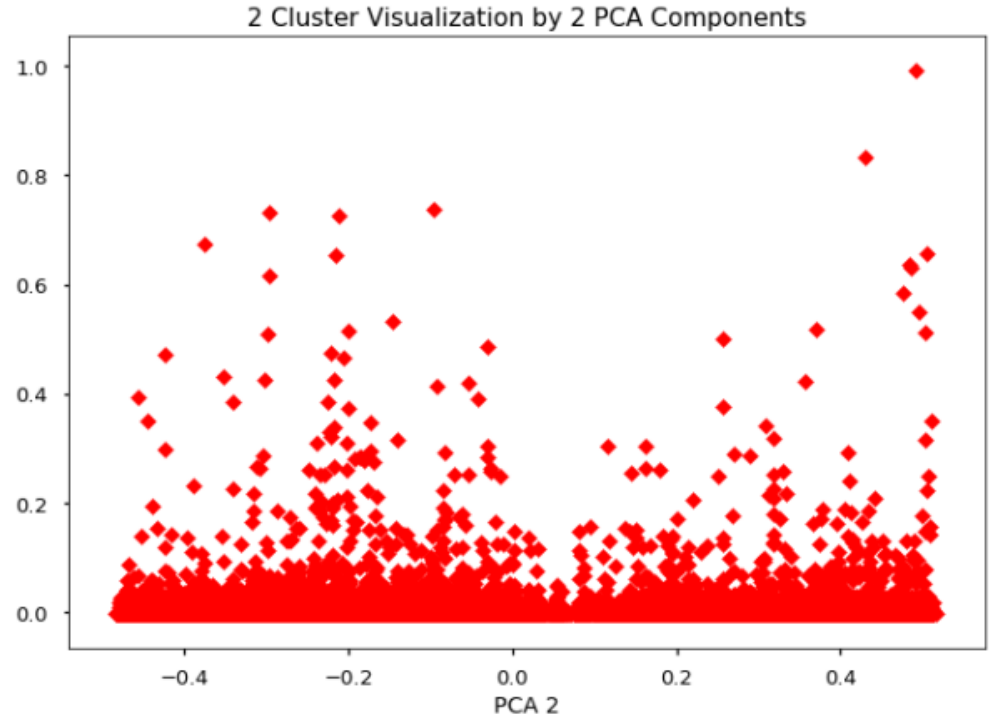
3

비지도 학습모델

이상 행위 탐지

- 다양한 방법을 이용해 K-means 알고리즘 테스트 수행

cluster	0	1	2	3
label				
0	23295	16475	29193	19604
1	1	3	6	1996



(정상 + 악성)데이터를 이용

악성(label=1)은 많이 분류되었지만,

오탐이 많이 발생

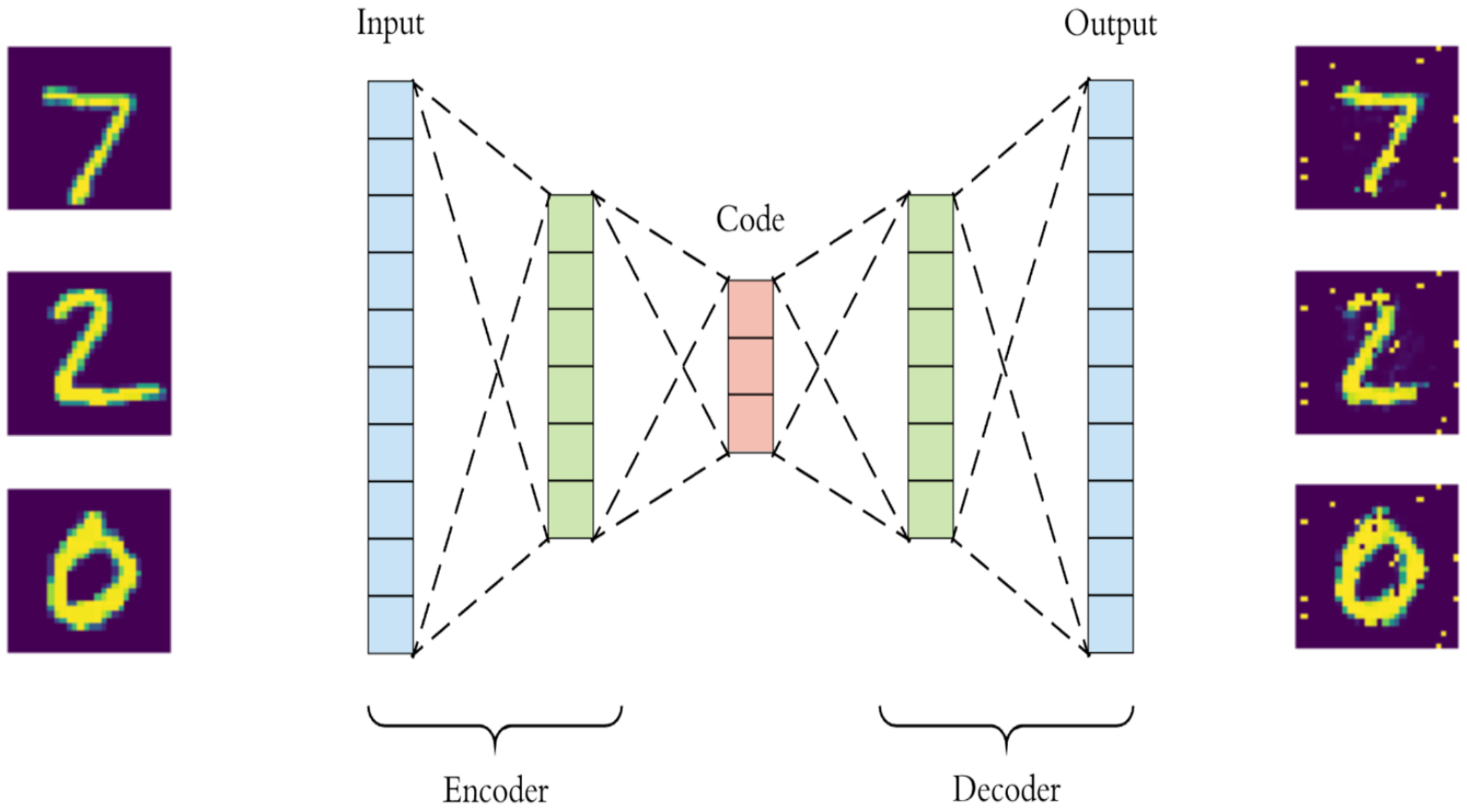
K-means 사용 불가 판단

3

비지도 학습모델

이상 행위 탐지

- 오토 인코더(Auto Encoder)
  - ➢ INPUT이 인코딩과 디코딩을 거쳤을 때, OUTPUT이 INPUT과 유사한 결과를 보여줌



3

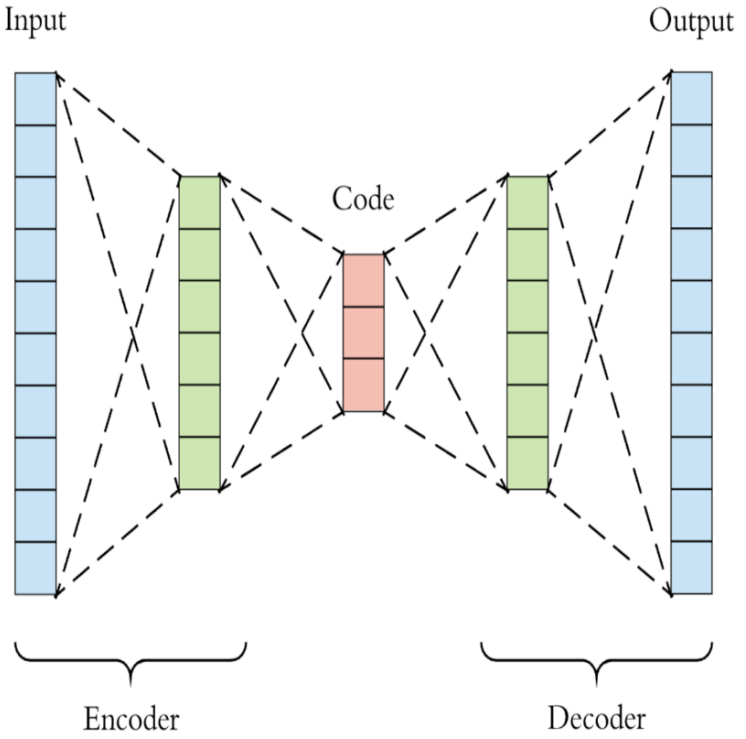
비지도 학습모델

이상 행위 탐지

- 정상 데이터만을 이용해서 학습
  - 정상 데이터를 넣으면 디코더 후에 정상과 유사한 값이 나옴
  - 정상 이외의 데이터를 넣으면 디코더 상에서 이상한 결과 값이 나옴

(정상)

(이상)



(정상)

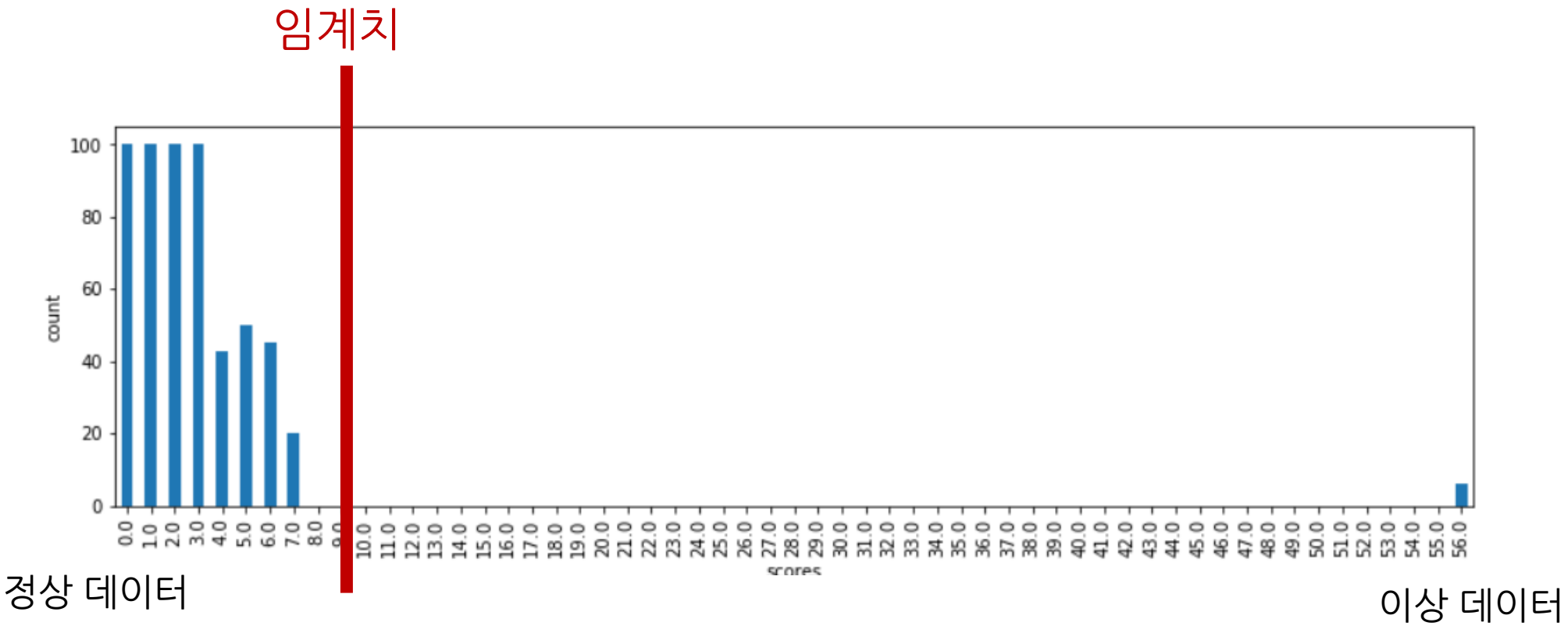
(이상)

3

비지도 학습모델

이상 행위 탐지

- 이상 데이터를 탐지하기 위해 임계치를 설정
  - ➢ 정상 데이터와 이상 데이터 간의 디코더 결과를 확인하여 임계치를 설정



3

비지도 학습모델

이상 행위 탐지

- 오토 인코더 결과

▶ 임계치를 32로 준 경우(A network)

```
a = data[data['result']>32]
a['ip.src'].value_counts()

205.174.165.73    736
Name: ip.src, dtype: int64
```

A	B	C	D
source_ip	destination_ip	source_port	destination_port
205.174.165.73	192.168.10.14	8080	51583
205.174.165.73	192.168.10.14	8080	51584
205.174.165.73	192.168.10.14	8080	51585
205.174.165.73	192.168.10.14	8080	51586
205.174.165.73	192.168.10.14	8080	51587
205.174.165.73	192.168.10.14	8080	51588
205.174.165.73	192.168.10.14	8080	51592
205.174.165.73	192.168.10.14	8080	51593
205.174.165.73	192.168.10.14	8080	51594

▶ 임계치를 7로 준 경우(B network)

```
3  a = data[data['result']>7]
4  a['ip.src'].value_counts()

41.216.186.89    2
207.46.13.202    1
66.249.71.34     1
Name: ip.src, dtype: int64
```

A	B	C	D
source_ip	destination_ip	source_port	destination_port
66.249.71.34	10.89.32.19	60604	8080
207.46.13.202	10.89.32.19	7048	80

3

비지도 학습모델

공격 유형 탐지

- 위 결과를 바탕으로 아래 Feature를 이용하여 공격 유형 탐지

No	Feature	Description	Attack Type
1	tot_fw_pk	위 내용과 동일	DoS DDoS Portscan Brute force
2	fw_iat_tot		
3	fw_iat_avg		
4	fw_iat_std		
5	fw_iat_max		
6	fw_iat_min		
7	fw_pkt_s		
8	ratio		
9	is_in_bot	http.user_agent에 bot에 관련된 문자열이 포함되어 있는지 여부	Bot
10	Is_in_ares	http.server에 bot과 관련된 문자열이 포함되는지 여부	
11	is_in_xss	http.request.uri에 XSS 공격과 관련된 문자열이 포함되어 있는지 여부	XSS
12	is_in_sql	httprequest.uri에 SQL injection 공격과 관련된 문자열이 포함되어 있는지 여부	SQL injection
13	check_rdp	tcp.dstport로 3389 포트를 사용하는지 여부	RDP attack

4

결론

본선 결과

- Network A 에서는 1346개의 악성 패킷을 발견(1개 공격유형)
- Network B 에서는 3개의 악성 패킷을 발견(2개 공격유형)

1위	와이앤슈	14:26:18 43.01%	14:45:21 38.74%	15:41:43 43%	15:55:04 43.01%	16:27:12 78.26%	14:26:18 67.59%	14:45:21 90%	15:41:43 84.44%	15:55:04 90%	16:27:12 90%
2위		14:27:39 89.51%					14:27:39 58.18%				
3위		14:27:35 38.49%	15:01:16 31.99%	15:13:39 39.66%	15:51:15 9.03%	15:56:02 39.66%	14:27:35 80%	15:01:16 80%	15:13:39 80%	15:51:15 80%	15:56:02 80%
4위		14:33:17 0%	15:12:30 0%	16:06:49 0%			14:33:17 0%	15:12:30 0%	16:06:49 0%		

4

결론

활용 방안

- 만약 우리가 네트워크 이상 탐지 솔루션을 구축한다면?

1. **실제 네트워크 상에서 학습을 위한 정상과 악성 패킷을 구분하기가 쉽지 않음**

   -> IDS, IPS 등을 활용하여 정상 패킷만 수집할 수 있어야 함

2. **Pcap파일을 통해 더 많은 피쳐를 추출 가능**

   -> TCP Flag 등을 이용 가능

3. **이상 탐지가 되었지만 회사 정책상 필요한 경우가 있음**

   -> 전문가 및 담당자 파악을 통해 예외 처리 필요

Thank you

대회를 마치며...